

Bayesian methods for analysing ringing data

S. P. BROOKS¹, E. A. CATCHPOLE², B. J. T. MORGAN³ & M. P. HARRIS⁴, ¹University of Cambridge, UK, ²University of New South Wales at ADFA, Australia, ³University of Kent at Canterbury, UK and ⁴Centre for Ecology and Hydrology, UK

ABSTRACT *A major recent development in statistics has been the use of fast computational methods of Markov chain Monte Carlo. These procedures allow Bayesian methods to be used in quite complex modelling situations. In this paper, we shall use a range of real data examples involving lapwings, shags, teal, dippers, and herring gulls, to illustrate the power and range of Bayesian techniques. The topics include: prior sensitivity; the use of reversible-jump MCMC for constructing model probabilities and comparing models, with particular reference to models with random effects; model-averaging; and the construction of Bayesian measures of goodness-of-fit. Throughout, there will be discussion of the practical aspects of the work—for instance explaining when and when not to use the BUGS package.*

1 Introduction

The paper by Racine *et al.* (1986) provided a clear signal that Bayesian methods were no longer a branch of academic statistics, as in, for example, Lindley (1965). The subsequent arrival of Markov chain Monte Carlo (MCMC) procedures (see for example, Smith & Gelfand (1992)), coupled with greatly increasing computer speeds, have resulted in the ‘Bayesian boom’, described by Malakoff (1999). Although the new methodology is now established in areas such as medical statistics (see Gilks *et al.*, 1996), Anderson & Burnham (1999) comment that ‘these methods remain relatively unknown to biologists’. There has, however, been some Bayesian analysis of data from previously marked animals. The papers by Castledine (1981), Smith (1988), Underhill (1990), Bolfarine *et al.* (1992), George & Robert (1992), Garthwaite *et al.* (1995), Madigan & York (1997) and Lee & Chen (2000) have as their primary emphasis the estimation of population size, while Janz (1980), Freeman (1990), and Link & Cam (2000), consider estimation of survival probabilities.

Correspondence: S. P. Brooks, Statistical Laboratory, University of Cambridge, Cambridge, CB3 0WB, UK. E-mail: S.P.Brooks@statslab.cam.ac.uk.

Dupuis (1995) and King & Brooks (2001) provide procedures for multi-site models, in which it is necessary to estimate probabilities of movement between sites.

In this paper, we provide several examples of the use of MCMC for analysing mark recovery and recapture data. We start, in Section 2, with an application to mark-recapture-recovery data. In this example we consider only one model; in practice we need to consider a range of alternative models for any data set. In Section 3 we present several alternative models for a mark-recovery data set, and demonstrate how posterior model probabilities can be estimated. In this case, model averaging is not needed; in Section 4, however, we use model averaging for a mark-recapture data set. In Section 5 we show how Bayesian methods are easily used to incorporate random effects, and in Section 6 we use reversible jump Markov chain Monte Carlo to produce posterior model probabilities in an example with individual covariates. In the general discussion of Section 7, we cover prior sensitivity and computational aspects, which both relate to using Bayesian methods in practice.

Throughout, the emphasis is on applied modelling, and we do not provide much technical detail. We do not discuss how prior distributions may be elicited, using the opinions of experts (see O'Hagan, 1998 for a discussion), but emphasize, when appropriate, whether results are sensitive to the choice of prior distributions.

This paper uses examples to illustrate the operation of modern Bayesian procedures. Many of the examples are historical, in that they were collected, and originally analysed, several years ago. However, we shall also consider current data sets on the survival of British birds, which are of topical interest, and each year sees the data extending in time. The first example, which we shall now consider, is of this form.

2 The survival of shags (*Phalacrocorax aristotelis*)

Catchpole *et al.* (1998) analyse nine years of mark-recapture-recovery data on shags ringed on the Isle of May in the Firth of Forth, Scotland. In total, 8637 birds were marked in the study, comprising 1087 breeding adults and 7550 pulli. The model they chose to describe the data has year-dependent first-year survival probabilities, $\{\phi_{1,i}\}$; a separate survival probability ϕ_{imm} for birds aged 1 and 2 years; a survival probability ϕ_a associated with all birds over the age of 3; age-dependent recapture probabilities p_i , $0 \leq i \leq 2$, for birds aged 0–2; a year-dependent adult recapture probability, $\{p_{a,i}\}$; and a year-dependent recovery probability, $\{\lambda_i\}$. This results in a model with 32 identifiable parameters in total. The model was fitted to the data using classical maximum-likelihood (ML). Although the model-fitting could be accomplished by using MARK (White & Burnham, 1999), Catchpole *et al.* (1998) maximized directly a likelihood that was specified via a set of sufficient matrices, a point to which we shall return later in the paper. We note that there have been very many other papers dealing with mark-recapture-recovery data, a particularly important and influential one being by Burnham (1993). In Table 1 we give the ML estimates of the parameters, along with estimated standard errors, which result from inverting a standard, second-order difference approximation to the Hessian matrix, evaluated at the ML estimate.

As is now well known, a Bayesian analysis updates an experimenter's beliefs about a parameter vector θ , originally expressed through a prior distribution $p(\theta)$, after the data \mathbf{x} have been observed. If we write $\pi(\theta|\mathbf{x})$ as the posterior distribution for θ , then Bayes' theorem tells us simply that

$$\pi(\theta|\mathbf{x}) \propto L(\mathbf{x}|\theta)p(\theta)$$

TABLE 1. Maximum likelihood parameter estimates, together with estimated asymptotic standard errors for the shags. The mean first year survival probability is estimated by $\hat{\phi}_1 = \Sigma \phi_{1,i}/9$. From Catchpole *et al.* (1998)

Survival			Recapture			Recovery		
	Estimate	St.Err.		Estimate	St.Err.		Estimate	St.Err.
$\phi_{1,1}$	0.325	0.037	p_1	0.0008	0.0005	λ_1	0.146	0.016
$\phi_{1,2}$	0.439	0.035	p_2	0.011	0.002	λ_2	0.116	0.012
$\phi_{1,3}$	0.193	0.035	p_3	0.034	0.005	λ_3	0.121	0.014
$\phi_{1,4}$	0.732	0.064	$p_{a,1}$	0.363	0.052	λ_4	0.093	0.018
$\phi_{1,5}$	0.441	0.051	$p_{a,2}$	0.462	0.034	λ_5	0.117	0.014
$\phi_{1,6}$	0.613	0.088	$p_{a,3}$	0.060	0.014	λ_6	0.046	0.010
$\phi_{1,7}$	0.397	0.076	$p_{a,4}$	0.012	0.005	λ_7	0.044	0.008
$\phi_{1,8}$	0.231	0.114	$p_{a,5}$	0.081	0.011	λ_8	0.097	0.014
$\phi_{1,9}$	0.767	0.083	$p_{a,6}$	0.276	0.022	λ_9	0.070	0.016
$\bar{\phi}_1$	0.460	0.023	$p_{a,7}$	0.018	0.004			
ϕ_{imm}	0.698	0.021	$p_{a,8}$	0.196	0.017			
ϕ_a	0.866	0.012	$p_{a,9}$	0.135	0.013			

where L denotes the likelihood. MCMC allows us to simulate a set of values from $\pi(\theta|\mathbf{x})$, so that Bayesian inference may be based upon posterior means, for example. MCMC algorithms typically proceed by updating each parameter, conditioning upon the value of the others. Thus, knowledge of the dependencies between the different parameters in the model is vital. These can be very efficiently expressed in the form of a directed acyclic graph (DAG: Gilks *et al.*, 1996, p. 25). A DAG consists of a series of nodes and interconnecting edges. Square nodes represent variables whose values are known (e.g. data and prior parameters) and circles denote unknown variables, typically model parameters that need to be estimated. An edge between two nodes denotes a direct relationship between the two corresponding variables. As well as providing an efficient mechanism for representing dependencies within the model, the DAG is also a very useful tool for developing and communicating model structure. For illustration, we provide in Fig. 1 the DAG for the model used to analyse the shags data. If we take independent uniform prior distributions for all of the parameters, then we obtain the results of Table 2. The agreement between the classical and Bayesian results is striking.

The classical analysis requires the formulation of the likelihood as a surface in 32 dimensions, and then numerical maximization of that surface. The Bayesian analysis is described in detail in Brooks *et al.* (2002), making use of procedures outlined in Brooks *et al.* (2000a). As we see from the above statement of Bayes theorem, it is also necessary here to construct the likelihood; and the existence of sufficient matrices, mentioned above, also results in a large computational saving for the Bayesian analysis. The Bayesian analysis requires no numerical optimization; instead it estimates the mean (or any other parameter of interest), through repeated simulations of a Markov chain with stationary distribution π . If the prior distributions are independent, $\text{Unif}(0,1)$, over all parameters, so that $p(\theta) = 1$, then maximizing the likelihood is equivalent to finding the mode of the posterior distribution. This explains the close agreement between Tables 1 and 2. In fact this agreement can be regarded as a vindication of the asymptotic assumptions required in the classical analysis.

We discuss the role of the prior distribution, $p(\theta)$, later. We return to this example in Sections 5 and 6.

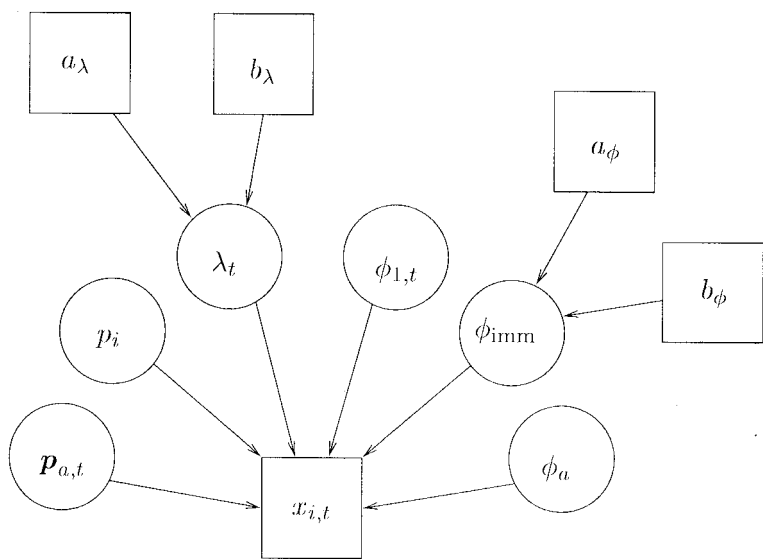


FIG. 1. Graphical representation of model for shags data: $\lambda \sim \text{Beta}(a_\lambda, b_\lambda)$, $\phi_{\text{imm}} \sim \text{Beta}(a_\phi, b_\phi)$, etc. ... ϕ denotes a survival probability, λ a recovery probability and p a recapture probability; x denotes the data. For clarity, only the prior parameters for λ and ϕ_{imm} are included.

TABLE 2. The result of a Bayesian analysis of the shags data corresponding to the classical analysis of Table 1, with independent $U(0,1)$ prior distributions for each parameter, showing posterior means and standard deviations (SD)

Survival			Recapture			Recovery		
Mean	SD		Mean	SD		Mean	SD	
$\phi_{1,1}$	0.328	0.038	p_0	0.001	0.001	λ_1	0.148	0.016
$\phi_{1,2}$	0.439	0.035	p_1	0.012	0.002	λ_2	0.117	0.012
$\phi_{1,3}$	0.198	0.035	p_2	0.034	0.005	λ_3	0.122	0.014
$\phi_{1,4}$	0.726	0.061	$p_{a,1}$	0.368	0.052	λ_4	0.095	0.018
$\phi_{1,5}$	0.441	0.051	$p_{a,2}$	0.464	0.034	λ_5	0.118	0.014
$\phi_{1,6}$	0.610	0.083	$p_{a,3}$	0.064	0.014	λ_6	0.047	0.010
$\phi_{1,7}$	0.403	0.076	$p_{a,4}$	0.014	0.006	λ_7	0.046	0.008
$\phi_{1,8}$	0.259	0.102	$p_{a,5}$	0.083	0.011	λ_8	0.101	0.014
$\phi_{1,9}$	0.732	0.097	$p_{a,6}$	0.279	0.022	λ_9	0.069	0.015
$\bar{\phi}_1$	0.460	0.023	$p_{a,7}$	0.019	0.004			
ϕ_{imm}	0.696	0.021	$p_{a,8}$	0.199	0.017			
ϕ_a	0.864	0.013	$p_{a,9}$	0.138	0.013			

3 Blue-winged teal (*Anas discors*)

Brownie *et al.* (1985) describe a well-known ring-recovery study of blue-winged teal, ringed as young in Saskatchewan, and in some cases later reported dead. The data are given in Table 3. This study also involved the ringing of teal ringed as adults, but we do not consider those data here. A Bayesian analysis of these data is provided by Brooks *et al.* (2000a) who consider four possible alternative models. In these models we use the subscript i to denote age dependence, i.e. a separate parameter for each age of the bird, and t to denote year dependence. Thus, for example, the model $\{\phi_{1,t}, \phi_i\}/\lambda_t$ has year-dependent survival probabilities in the

TABLE 3. Recoveries of blue-winged teal marked as young from 1961 to 1973. Data from Brownie *et al.* (1985)

Year of ringing	Number ringed	Year of recovery (1961+)											
		1	2	3	4	5	6	7	8	9	10	11	12
1961	910	6	2	1	1	0	2	1	0	0	0	0	0
1962	1157		11	5	6	1	1	1	1	0	0	0	1
1963	1394			19	4	4	4	0	0	1	1	0	0
1964	3554				65	25	8	4	2	4	4	1	0
1965	4849					65	17	2	1	6	2	3	1
1966	2555						52	9	8	3	4	2	1
1967	305							3	1	0	1	0	0

first year of life, with a separate, but non-year-dependent, survival probability for each subsequent year of life, and recovery probabilities that are year- but not age-dependent. The other models considered are submodels of this: for example the model $\{\phi_1, \phi_a\}/\lambda_t$ has first year survival probabilities that are constant over time, and a single constant annual survival probability ϕ_a for all birds aged ≥ 1 .

The basic expression for the Bayes Theorem given above may be elaborated to include models. Thus, if m is a parameter that indexes the various models we can write

$$\pi(\theta_m, m | \mathbf{x}) \propto L_m(\mathbf{x} | \theta_m) p(\theta_m | m) p(m) \tag{1}$$

Here, θ_m is the vector of parameter values under model m , $p(m)$ is the prior distribution over models, $p(\theta_m | m)$ is the prior distribution for θ_m under model m , and $L_m(\mathbf{x} | \theta_m)$ is the likelihood for model m . For discussion of model probabilities, see Carlin & Louis (1996, Section 2.4.2). What we obtain from (1) is the joint posterior distribution of θ_m and m . If we seek posterior model probabilities then these can be obtained from the posterior marginal:

$$\pi(m | \mathbf{x}) \propto p(m) \int L_m(\mathbf{x} | \theta_m) p(\theta_m | m) d\theta$$

This integrand is simply the expected value of $L_m(\mathbf{x} | \theta_m)$ with respect to the prior and can be estimated by taking the sample mean of the likelihood, given a series of θ_m values drawn from the prior. Although, in this case, the method appears to work well, in many others prohibitively large sample sizes are required in order to obtain accurate estimates. A wide range of alternative methods are available in such cases, see Gamerman (1997, Section 7.2), for example. We discuss this issue further in Section 7.2. We give in Table 4 the posterior model probabilities for the four models.

For each model, the prior distribution for the different model parameters were taken to be independent, $\text{Unif}(0, 1)$. We may therefore conclude from Table 4 that for this choice of prior distributions, of the four models considered, $\{\phi_1, \phi_a\}/\lambda_t$ provides the best description of the data. This was also the conclusion from the classical analysis of these data by Freeman & Morgan (1992), and their ML estimates of the model parameters agree well with the posterior means, as in Section 3. A common defect of classical analyses is that results are often presented for a ‘best’ model and do not reflect the stages taken in the model-selection process. The availability of posterior model probabilities, following a Bayesian analysis, means that model-averaging may readily take place. However, we can see that, in

TABLE 4. Posterior means and standard deviations (SD) under independent uniform priors, for four models fitted to the teal data set, from a sample of 10 000 observations from the posterior distribution, together with the posterior model probabilities, $\pi(m|\mathbf{x})$, associated with each model. From Brooks *et al.* (2000a)

ϕ_i/λ			$\{\phi_{1,i}, \phi_i\}/\lambda$		$\{\phi_{1,i}, \phi_i\}/\lambda_i$		$\{\phi_{1,i}, \phi_a\}/\lambda_i$	
Mean		SD	Mean	SD	Mean	SD	Mean	SD
$\phi_{1,1}$			0.531	0.126	0.525	0.248		
$\phi_{1,2}$			0.588	0.089	0.531	0.114		
$\phi_{1,3}$			0.431	0.081	0.321	0.083		
$\phi_{1,4}$			0.428	0.047	0.484	0.070		
$\phi_{1,5}$			0.396	0.047	0.259	0.055		
$\phi_{1,6}$			0.352	0.052	0.264	0.077		
$\phi_{1,7}$			0.459	0.179	0.158	0.108		
ϕ_1	0.420	0.026					0.403	0.039
ϕ_2	0.604	0.040	0.602	0.038	0.445	0.071		
ϕ_3	0.694	0.047	0.690	0.049	0.538	0.087		
ϕ_4	0.640	0.062	0.636	0.059	0.459	0.094		
ϕ_5	0.690	0.073	0.682	0.070	0.578	0.100		
ϕ_6	0.624	0.093	0.616	0.091	0.545	0.108		
ϕ_7	0.410	0.133	0.397	0.130	0.372	0.134		
ϕ_8	0.494	0.207	0.474	0.206	0.468	0.202		
ϕ_9	0.729	0.217	0.722	0.219	0.749	0.205		
ϕ_{10}	0.688	0.232	0.695	0.230	0.691	0.233		
ϕ_{11}	0.348	0.240	0.352	0.240	0.355	0.240		
ϕ_{12}	0.493	0.285	0.511	0.285	0.504	0.288		
ϕ_a							0.598	0.053
λ_1					0.058	0.137	0.013	0.005
λ_2					0.017	0.008	0.017	0.005
λ_3					0.018	0.005	0.024	0.005
λ_4					0.035	0.006	0.031	0.004
λ_5					0.019	0.003	0.026	0.003
λ_6					0.028	0.004	0.031	0.004
λ_7					0.019	0.005	0.016	0.004
λ_8					0.056	0.022	0.031	0.008
λ_9					0.078	0.040	0.035	0.011
λ_{10}					0.140	0.096	0.051	0.020
λ_{11}					0.149	0.123	0.048	0.025
λ_{12}					0.196	0.181	0.047	0.034
λ	0.026	0.001	0.026	0.001				
$\pi(m \mathbf{x})$	0.009		0.054		0.000		0.937	

this instance, due to the very high probability associated with model $\{\phi_1, \phi_a\}/\lambda_i$, model-averaging is not necessary in this instance. We provide an example of model-averaging in the next section. For further discussion of model-averaging, from differing perspectives, see Burnham & Anderson (1998, Ch. 4), Buckland *et al.* (1997), Chatfield (1995), Di Ciccio *et al.* (1997) and Raftery *et al.* (1997).

Although, for the family of models considered, $\{\phi_1, \phi_a\}/\lambda_i$ is the preferred fitted model, it might still provide a poor fit to the data; we need to consider a measure of absolute fit as well as the posterior model probabilities, which gauge relative fit. One way to do this is by calculating Bayesian p -values, as discussed in Vounatsou & Smith (1995) and Brooks *et al.* (2000a). To compare the data, \mathbf{x} , with model-generated expected values, \mathbf{e} , we form the discrepancy measure,

$$D(\mathbf{x}; \theta) = \sum_j (\sqrt{x_j} - \sqrt{e_j})^2$$

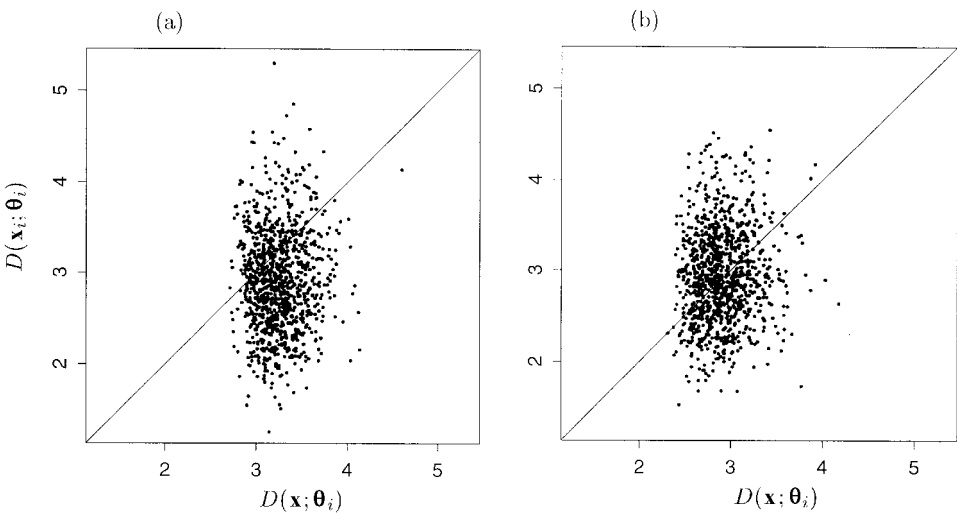


FIG. 2. Discrepancy plots for the models (a) $\{\phi_{1,i}, \phi_i\}/\lambda_i$; and (b) $\{\phi_1, \phi_a\}/\lambda_a$ for the teal data. Corresponding Bayesian p -values are 0.28 and 0.50, respectively. From Brooks *et al.* (2000a).

We take a random sample $\{\theta_i, 1 \leq i \leq n\}$ from the posterior distribution of θ , which is easily done using an MCMC sampler. For each θ_i we simulate a new data set, \mathbf{x}_i , and plot $D(\mathbf{x}_i; \theta_i)$ against $D(\mathbf{x}; \theta_i)$. The Bayesian p -value is the proportion of times that $D(\mathbf{x}_i; \theta_i)$ is greater than $D(\mathbf{x}; \theta_i)$: p -values close to 0.5 represent a good fit to the data.

The discrepancy plots for the two models with highest posterior probability are shown in Fig. 2. The better fit of model $\{\phi_1, \phi_a\}/\lambda_a$ to the data is reflected both in the p -values (0.50, as compared with 0.28 for model $\{\phi_{1,i}, \phi_i\}/\lambda_i$), and in the smaller values of $D(\mathbf{x}; \theta_i)$ evident in Fig. 2.

As discussed by Brooks *et al.* (2000a), there are alternative formulations of Bayesian p -values and, as with model probabilities, they depend on the prior distribution chosen.

4 Dippers (*Cinclus cinclus*)

We now turn to a capture-recapture data set, analysed by Lebreton *et al.* (1992), and given in Table 5.

TABLE 5. Capture-recapture data for European Dippers banded in 1981–86, taken from Lebreton *et al.* (1992). Here we have pooled the data originally given separately for males and females

Year of release	Number released	Year of Recapture (1981+)					
		1	2	3	4	5	6
1981	22	11	2	0	0	0	0
1982	60		24	1	0	0	0
1983	78			34	2	0	0
1984	80				45	1	2
1985	88					51	0
1986	98						52

We consider four models: ϕ_i/p_i ; ϕ_i/p ; ϕ/p ; and $\{\phi_n, \phi_f\}/p$. This last model has two survival rates, with a possible change due to a flood in 1983. For the ordering of the models given above, and independent $\text{Unif}(0, 1)$ prior distributions for each model parameter, the corresponding posterior model probabilities are 0.000, 0.000, 0.205 and 0.795. In this case there are two reasonable models. The favoured model, $\{\phi_n, \phi_f\}/p$, which is the one chosen by a classical analysis, has a Bayesian p -value of 0.125, as opposed to 0.069 for the ϕ/p model. If interest focuses on the survival probability, ϕ_n , say, of birds in non-flood years, then for the two models we obtain the respective posterior means (SDs), 0.561 (0.025) and 0.609 (0.031). Using model averaging, with respective probabilities of 0.205 and 0.795, we obtain the estimate, $\hat{\phi}_n = 0.599$ (0.038). By accounting for model-uncertainty, model-averaging results in a suitably inflated estimate of the standard deviation.

5 Random effects

We might wish to include a variety of different random effects in models for avian survival. An empirical Bayes approach is outlined in White & Burnham (1999) and is available in the program MARK—see Burnham (2001). It is quite simple to include random effects in a full Bayesian analysis, and to illustrate this, we start with a re-analysis of the shag data from Section 2.

5.1 Random effects for shags

A feature of Table 1 is that if one is primarily interested in estimating survival, there are many nuisance parameters, estimated as fixed effects, namely all the recapture and recovery probabilities. Additionally, we might want to use a random effect to describe the time-variation in the $\{\phi_{1,t}\}$. We show in Fig. 3 how we incorporate a random effect.

The model adopted here is, for example for the $\{\lambda_t\}$,

$$\text{logit } \lambda_t = \mu_\lambda + \varepsilon_t \tag{2}$$

We follow Zeger & Karim (1991), and introduce random effects on the logistic scale. It would also be possible to introduce random effects directly on the probability scale, for example through beta distributions. However, we expect both approaches would lead to similar conclusions (cf. Williams, 1988). Here μ_λ is a basic model parameter that is given a suitably vague prior distribution, say, $N(0, \sigma_\mu^2)$ with large σ_μ . The model becomes a random effects model by letting the $\{\varepsilon_t\}$ be independent, $N(0, \sigma_\varepsilon^2)$, and giving σ_ε^2 a suitable prior. We use an inverse gamma prior distribution with parameters a_ε and b_ε , which we denote by $\Gamma^{-1}(a_\varepsilon, b_\varepsilon)$. This is the customary conjugate prior to use in this case, but many alternatives are, of course, possible. Thus, precisely as in Zeger & Karim (1991), the $\{\varepsilon_t\}$ and σ_ε^2 are basic model parameters and treated equally in the posterior distribution. In contrast, σ_μ , a_ε and b_ε are prior parameters, which are fixed in order to specify prior beliefs.

The results of having random effects for $\{\phi_{1,t}\}$, $\{p_{a,t}\}$ and $\{\lambda_t\}$ are shown in Table 6. Comparing with Table 1, we see that the effect on the remaining parameters, such as ϕ_a , is minimal. What Table 6 tells us is how we may describe the observed yearly variation in the ϕ_1 , p_a and λ parameters through random effects. Table 6 gives the means and standard deviations of the marginal distributions of the parameters considered there. The individual parameters, $\{\phi_{1,t}\}$, $\{p_{a,t}\}$ and $\{\lambda_t\}$ also have poster-

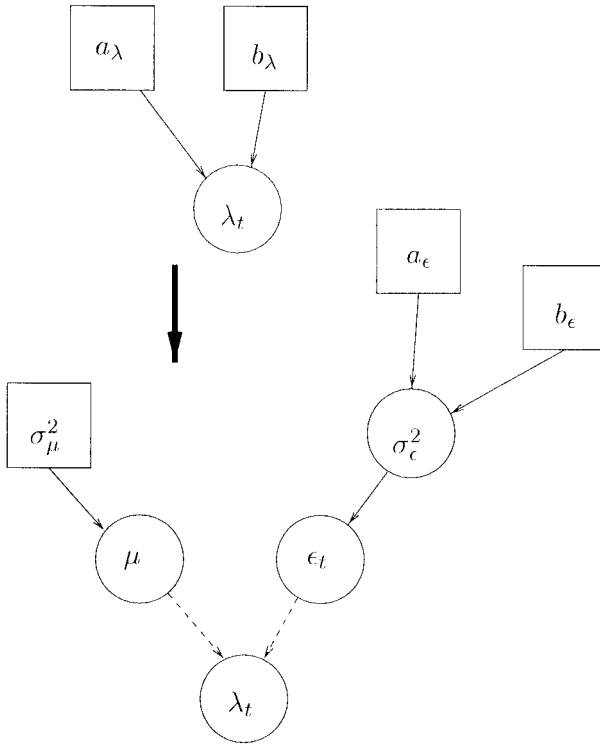


FIG. 3. A diagrammatic representation of the changes required to convert a fixed effects model for time-varying recovery probabilities (upper diagram) into a random effects model (lower) for the shags data. Here, circled parameters are simulated, while boxed ones are fixed in advance, and hence represent prior knowledge. Solid arrows indicate probabilistic dependencies, while dashed arrows show deterministic relationships. The random effects model is $\text{logit } \lambda_t = \mu + \epsilon_t$, $\mu \sim N(0, \sigma_\mu^2)$, $\epsilon_t \sim N(0, \sigma_\epsilon^2)$ and $\sigma_\epsilon^2 \sim \Gamma^{-1}(a_\epsilon, b_\epsilon)$.

TABLE 6. The results of fitting the random effects model of Section 5.1 to the shag data. Priors were as follows: for ϕ_{imm} , ϕ_a , p_1 , p_2 and p_3 , $\text{Unif}(0,1)$; for μ_{ϕ_1} , μ_{p_a} and μ_λ , $N(0,100)$; for $\sigma_{\phi_1}^2$, $\sigma_{p_a}^2$ and σ_λ^2 , $\Gamma^{-1}(1,1)$.

	Mean	SD		Mean	SD
μ_{ϕ_1}	-0.094	0.517	ϕ_{imm}	0.697	0.021
$\sigma_{\phi_1}^2$	1.225	0.534	ϕ_a	0.864	0.012
μ_{p_a}	-1.973	0.714	p_1	0.001	0.001
$\sigma_{p_a}^2$	2.074	0.873	p_2	0.012	0.003
μ_λ	-2.319	0.242	p_3	0.034	0.006
σ_λ^2	0.695	0.280			

ior distributions and, for example, their expectations may be of interest. We might also want to consider how well a model with just the random effects of Table 6 might describe the data. This could be done using the deviance information criterion (DIC) of Spiegelhalter *et al.* (2001), see also Barry *et al.* (2002).

5.2 Random effects for lapwings

Catchpole *et al.* (1999) provide a classical analysis of recovery data on British lapwings (*Vanellus vanellus*). They fitted a model with a declining recovery prob-

TABLE 7. Recovery data for British Lapwings ringed as chicks during the years 1970-1992. Data from Catchpole *et al.* (1999)

Year of ringing	Number ringed	Year of recovery																						
		71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93
1970	1963	8	3	2	0	0	0	1	2	3	0	0	0	0	0	1	0	0	0	0	0	0	0	0
1971	2463		4	1	1	2	2	1	3	2	0	2	0	0	0	0	0	0	0	0	0	0	0	0
1972	3092			7	2	2	2	5	1	5	2	1	3	1	1	1	2	0	0	0	0	0	0	0
1973	3442				15	1	1	1	2	3	2	0	1	1	1	1	0	0	0	0	0	0	0	0
1974	3945					12	4	4	7	4	1	1	1	1	0	1	2	1	1	0	0	0	0	0
1975	2564						12	3	0	3	1	1	1	2	0	3	0	0	0	0	0	0	0	0
1976	3304							11	3	5	1	3	4	2	3	0	1	0	1	1	0	0	0	0
1977	3478								13	6	0	5	4	2	1	2	4	0	0	0	1	0	0	0
1978	3165									18	5	2	2	0	5	3	0	0	0	1	0	0	0	0
1979	3351										7	4	7	1	4	3	3	0	1	0	0	1	0	1
1980	3487											12	8	2	2	4	5	3	0	1	2	0	0	1
1981	3939												26	2	3	3	3	4	1	0	0	0	0	0
1982	2881													11	0	6	3	0	1	2	0	0	1	0
1983	4119														19	7	7	4	0	0	2	1	0	0
1984	4036															13	3	2	0	0	0	0	1	0
1985	4867																23	2	5	2	0	3	3	2
1986	4769																	13	4	3	3	4	2	2
1987	5027																		14	2	1	2	3	2
1988	4804																			17	4	4	3	0
1989	3632																				11	4	2	4
1990	4170																					12	3	3
1991	4314																						9	4
1992	3480																							18

ability, and with the annual survival probabilities of first-year and adult birds regressed on measures of winter severity. Here, we reanalyse a subset of these data, shown in Table 7, chosen to simplify the model, so that the recovery probability parameter can be taken as constant. Details are given in Barry *et al.* (2002). We write ϕ_{ij} as the probability that a bird ringed in year i survives its j th year of life, given that it was alive at age $j - 1$, and consider the random effects model

$$\text{logit } \phi_{ij} = \psi_j + b_{i+j} + \varepsilon_{ij} \tag{3}$$

This is a mixed model: the $\{\psi_j\}$ are fixed-effect parameters measuring age dependency, $\{b_k\}$ are random year effects, and $\{\varepsilon_{ij}\}$ are error terms. We assume that $b_k \sim N(0, \sigma_b^2)$ and $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$. Thus, the model of (3) is a straightforward generalization of the model of (2), to include several fixed effects, reflecting general age variation, and two components of variance. An attractive feature of this formulation is the simple and natural way in which it can allow the inclusion of overdispersion, through σ_ε^2 . Cf. Lebreton *et al.* (1992), who account for overdispersion in the traditional way of scaling up standard errors by an overdispersion factor reflecting the lack of fit of the model to the data. The parameters given prior distributions in (3) are $\{\psi_j\}$, σ_b^2 and σ_ε^2 , which combine with a prior distribution for the recovery probability to complete the prior model specification.

The model (3) is readily extended to incorporate group covariates which vary over years, in common for all the birds of that year, to produce the mixed model

$$\text{logit } \phi_{ij} = \psi_j + \mathbf{x}_{ij}'\beta + b_{i+j} + \varepsilon_{ij} \tag{4}$$

where \mathbf{x}_{ij} denotes a vector of covariates.

For the data of Table 7, the first-year survival probability was found not to vary with weather. We consider the following model, involving just two age-classes for survival,

$$\begin{aligned}\text{logit } \phi_{i1} &= \psi_1 + b_{i+j} + \varepsilon_{ij} \\ \text{logit } \phi_{ij} &= \psi_a + \beta x_{i+j} + b_{i+j} + \varepsilon_{ij}, \quad \text{for } j > 1\end{aligned}\tag{5}$$

in which x_k denotes the number of days with lying snow in the third quarter of year k . We may fit this model, (a) with only ψ_1 , ψ_a and β and no random effects, (b) with added $\{b_k\}$ only, (c) with added $\{\varepsilon_{ij}\}$ only, and (d) with added $\{b_k\}$ and $\{\varepsilon_{ij}\}$. The resulting Bayesian p -values are: (a) 0.26, (b) 0.60, (c) 0.63, (d) 0.76. We conclude that the basic regression of adult annual survival on snow can be improved by the addition of either random yearly variation in addition to that explained by the regression, or random overdispersion, but that both effects do not appear to be necessary. For models (a) and (c), the posterior means and standard deviations for β , from (5), are, respectively, 0.45 (0.07) and 0.38 (0.09). As expected, the addition of overdispersion increases the estimate of standard deviation. What is also interesting is the change in the estimate of β , which may reflect influential data points whose presence is absorbed by the overdispersion.

The analysis of covariates presented here is rudimentary and illustrative only. In practice, of course, we have the situation faced by Catchpole *et al.* (1999), in which there are very many alternative possible weather covariates vying for possible inclusion in the model. Biological knowledge needs to be considered here. While one may devise ad hoc statistical approaches based on classical stepwise procedures, augmented by AIC (Catchpole *et al.*, 1999, 2000), we believe that modern Bayesian methods may provide an efficient, viable, alternative way of proceeding, and we describe these methods next.

6 Dealing with covariates

We saw in the previous section that we are often interested in the effects of covariates measuring aspects of the climate, for example, on survival. We are also interested in individual differences. Adding or subtracting single covariates to a model might well correspond to the different neighbouring models that RJMCMC is designed for. In this section, we provide an illustration of RJMCMC at work on the shag data, for a single individual covariate indicating sex. Female shags are usually mute, so that when adults are recaptured, it can be easy to sex them. The covariate takes one of three possible values, 0 (unknown), 1 (male) or 2 (female). The ‘unknown’ code refers primarily to birds ringed as chicks which were not later observed breeding. Thus, we can expect these birds to have smaller estimated survival and recapture probabilities. This has been found to be the case, and in the analysis below we focus on the effects of the male and female codes, 1 and 2 respectively, for each of the model parameters.

In Table 8 we give the posterior probabilities, evaluated using RJMCMC, for models based on $\{\phi_1, \phi_{\text{imm}}, \phi_a\} / \{p_1 = p_2 = p_3 = p, p_a\} / \lambda$, with a common probability p of recapture in the first three years of life, and no time dependence in any parameter. The models are binary coded to indicate which terms include a sex effect, e.g. 001/01/0 includes a sex effect for ϕ_a and p_a .

As an illustration, for model 001/00/0 we write ϕ_a , in self-explanatory notation, as

TABLE 8. Models specified according to whether (1) or not (0) a sex effect has been included in the parameters of the model for, in order, ϕ_{13} , ϕ_{imm} , ϕ_a , p , p_a and λ , and the corresponding posterior model probabilities

Model	Prob	Model	Prob	Model	Prob	Model	Prob
000/00/0	0.0519	101/10/0	0.0027	101/11/0	0.0000	110/10/1	0.0000
100/00/0	0.0003	011/10/0	0.0089	011/11/0	0.0000	001/10/1	0.0641
010/00/0	00008	111/10/0	0.0001	111/11/0	0.0000	101/10/1	0.0003
110/00/0	0.0000	000/01/0	0.0032	000/00/1	0.0062	011/10/1	0.0011
001/00/0	0.1500	100/01/0	0.0000	100/00/1	0.0000	000/01/1	0.0004
101/00/0	0.0008	010/01/0	0.0001	010/00/1	0.0001	010/01/1	0.0000
011/00/0	0.0025	001/01/0	0.0010	001/00/1	0.0200	001/01/1	0.0002
111/00/0	0.0000	101/01/0	0.0000	101/00/1	0.0001	011/01/1	0.0000
000/10/0	0.1658	011/01/0	0.0000	011/00/1	0.0004	000/11/1	0.0011
100/10/0	0.0008	000/11/0	0.0102	111/00/1	0.0000	100/11/1	0.0000
010/10/0	0.0029	100/11/0	0.0001	000/10/1	0.0193	010/11/1	0.0000
110/10/0	0.0000	010/11/0	0.0002	100/10/1	0.0001	001/11/1	0.0005
001/10/0	0.4801	001/11/0	0.0033	010/10/1	0.0003	101/11/1	0.0000

$\text{logit}(\phi_a) = \mu,$

for birds of unknown sex, and

$\text{logit} \phi_{a,M} = \mu + \tau_M$

$\text{logit} \phi_{a,F} = \mu + \tau_F$

for birds of unknown sex, where $\tau_M \sim N(0, \sigma_\tau^2)$, $\tau_F \sim N(0, \sigma_\tau^2)$ and $\sigma_\tau^2 \sim \Gamma^{-1}(a_\tau, b_\tau)$. The posterior means and standard deviations are as follows:

μ	1.1673	(0.2216)
τ_M	1.1067	(0.2705)
τ_F	0.6128	(0.2519)
$\tau = \tau_M = \tau_F$	0.9004	(0.2407)
σ^2	1.1164	(0.8764)

Note that τ_M and τ_F are thus random effects. It might be more sensible to regard them as fixed effects. We see from Table 8 that there is a high probability of a sex effect for both p and ϕ_a . Because of the way that the sex code was adjusted for

TABLE 9. Posterior probabilities of a sex effect in individual parameters for the shags data. These are obtained from summing the posterior model probabilities of Table 8 over all models that include a sex effect for that parameter

Parameter	Probability
ϕ_{13}	0.0053
ϕ_{imm}	0.0174
ϕ_a	0.7362
p	0.7618
p_a	0.0205
λ	0.1143

birds ringed as chicks, the result for ϕ_a is likely to be biased. However, the result for p is interesting, as the value for young males is larger than that for young females, which we would expect from the biology, as attempts at early breeding are far more common in males than in females, resulting in the higher value of p for males. When we just consider birds ringed as adults, for which there is relatively little change in the sex specification following marking, we obtain posterior means $\phi_{a,M} = 0.8912$, and $\phi_{a,F} = 0.8161$. Note that Catchpole *et al.* (1998), from data on birds ringed only as adults, found the ML estimate $\hat{\phi}_a = 0.843$ (0.016). In this case the probability of a sex effect for adult survival is 0.89. No previous study has found a sex difference in the survival of adult shags.

7 Discussion

In this paper we have emphasized the positive features of the modern Bayesian approach to analysing ringing data. Our enthusiasm derives more from being able to derive posterior model probabilities and perform model averaging, as well as to devise simply coherent integrated ways to consider variable selection, rather than being able to incorporate prior beliefs through informed priors. Undoubtedly, however, this latter aspect is attractive and important. In this section of the paper we cover some of the practical features that need to be considered before Bayesian analyses can take place. The first issue is one of how to proceed computationally.

7.1 Computing

The paper by Link & Cam (2000) demonstrates how the formidable BUGS package (see Spiegelhalter *et al.*, 1996a, b) may be easily used for basic Bayesian analyses of ringing data, including random effects models. Brooks *et al.* (2000a) provide a full discussion of the relative advantages and disadvantages of using BUGS. Implementing BUGS code requires some familiarity with the S-plus computing language (Venables & Ripley, 1999). However WinBUGS, the windows version of BUGS comes with some very nice and well-explained examples. Our experience is that it is not difficult to learn how to use BUGS. An illustration is provided by the BUGS code of the Appendix. The BUGS package will handle multinomial data. However, greater computational efficiency will result from individually tailored programs written in compiled languages such as FORTRAN. The shag models that we have considered in this paper require age-dependent structures that do not result in straightforward multinomial models and, for the shag modelling of this paper, it was necessary to use specific FORTRAN programs. The same is true of reversible jump MCMC (RJMCMC) methods (Green, 1995), which we believe have enormous potential—see Section 7.2. We have not yet carried out a formal efficiency comparison between classical and Bayesian analyses. We expect that relative performance will depend critically upon both models and data. However, in applications such as that of Section 6, the use of RJMCMC could result in the most efficient approach. We plan to make this kind of comparison in future work.

7.2 Model averaging and posterior model probabilities

The posterior distribution of (1) may be explored using RJMCMC, which allows us to move between models as well as to consider different values for model

parameters. For RJMCMC to work well in practice, we need to be able to move readily between the different states of the relevant Markov chain. This is generally very easy to achieve when we have a nested model structure and we move from model to model by simply adding or deleting parameters.

A primary advantage of RJMCMC is that it focuses on the posterior distribution to evaluate posterior model probabilities. This is in contrast to the standard formulation of posterior model probabilities, which involves averaging over a prior distribution. As found in Brooks *et al.* (2000a), this can sometimes result in an impractical procedure, which requires prohibitive amounts of computer time, as discussed in Section 3.

The performance of the standard approach to model probabilities depends on how similar the prior and likelihood are. If they are very different, most samples are from the mode of the prior, and we get very small likelihood values. Occasionally, we will get a draw from the tail of the prior, and then get a large contribution from the likelihood. If the prior and likelihood are similar, then this happens to a lesser degree—we learn about the likelihood much more quickly and estimate model probabilities more efficiently. The work of Section 6 was very easy to program as, in that case, the RJMCMC moves were very easy to add. For the work of Sections 3 and 4, the RJMCMC updates involved moves that added a larger number of variables to a model, which is far more difficult to do in an efficient way. In that case, we evaluated mode probabilities using the standard approach that worked well.

7.3 Prior sensitivity

Model-probabilities and Bayesian p -values can be sensitive to the choice of prior distributions and examples of this sensitivity are provided by Brooks *et al.* (2000a). A striking direct illustration of prior sensitivity is given by Brooks *et al.* (2000b), which we repeat here. In Table 10 we present a historical data set describing the recoveries of Herring gulls (*Larus argentatus*). The ϕ_i/λ model is fitted to these data by Vounatsou & Smith (1995). In the case of this extensive data set, with so few recoveries after 12 years of life, it is inevitable that a model with fully age-dependent annual survival probabilities will produce parameter estimates that reflect the prior distribution when available data are sparse. This is seen clearly from the results of Table 11. There is a detailed discussion of these results in Brooks *et al.* (2000b). In choosing between competing models for a data set, Brooks *et al.* (2000) argue in favour of selecting models that are not sensitive to the choice of prior. Shown in Fig. 4 are some of the results from fitting the $\{\phi_1, \phi_a\}/\lambda_t$ model to the teal data. In this case, except when an inappropriate prior distribution is chosen for the time-varying recovery probabilities, the data dominate the posterior distribution, so that changing the prior distributions for ϕ_1 and ϕ_a has little effect. Brooks *et al.* (2000a) speculate that, as a general rule, models for ring recovery data are more likely to be sensitive to prior choice than models for recapture data.

In the work of this paper we have used vague prior distributions. However it is simple to impose restrictions on prior distributions—for example, one might want to impose the restriction that first-year survival probabilities are less than adult survival probabilities. In a single study of this we found that the restriction was never violated, but in other applications such restrictions could be important. A further aspect of prior sensitivity is sensitivity relative to an informed prior, making use of expert beliefs. We have not considered that in this paper.

TABLE 10. Recoveries of Herring gulls *Larus argentatus*, marked as young. Data from Paynter (1966)

Year of ringing	Number ringed	Year of recovery (- 1933)																														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29		
1934	3 646	62	15	9	7	13	8	4	3	0	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1935	10 748	169	67	29	22	20	13	10	4	4	2	2	2	4	1	2	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
1936	6 665		92	64	27	16	19	13	6	3	4	2	1	2	1	2	1	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0
1937	4 652			70	21	15	9	12	8	2	4	0	2	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1938	2 983				29	25	7	3	2	4	2	2	2	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1939	3 000					72	28	13	4	5	3	3	3	1	3	2	3	3	0	3	1	0	0	0	0	1	0	0	0	0	0	0

TABLE 11. The influence of the prior distribution when the ϕ_i / λ model is fitted to the gull data. For λ we use Unif(0,1) priors throughout, while for the ϕ parameters we use either Unif(0,1) or Beta(4,2) or Beta(20,5). These priors are artificial and are chosen for illustration only. Note that for a Beta(a, b) distribution, the mean is $a/(a + b)$ and the variance is $ab/\{(a + b)^2(a + b + 1)\}$. The priors for all parameters are independent. From Brooks *et al.* (2000b)

Parameter	Unif(0,1)		Beta(4,2)		Beta(20,5)	
	mean	st.dev.	mean	st.dev.	mean	st.dev.
ϕ_1	0.550	0.015	0.551	0.015	0.556	0.015
ϕ_2	0.636	0.020	0.636	0.019	0.643	0.019
ϕ_3	0.739	0.022	0.740	0.023	0.745	0.022
ϕ_4	0.783	0.024	0.784	0.024	0.787	0.023
ϕ_5	0.682	0.031	0.684	0.031	0.696	0.029
ϕ_6	0.677	0.037	0.679	0.037	0.698	0.035
ϕ_7	0.736	0.044	0.734	0.043	0.754	0.038
ϕ_8	0.772	0.048	0.772	0.045	0.788	0.040
ϕ_9	0.790	0.050	0.789	0.049	0.801	0.043
ϕ_{10}	0.759	0.060	0.760	0.058	0.782	0.047
ϕ_{11}	0.768	0.067	0.768	0.063	0.793	0.051
ϕ_{12}	0.614	0.086	0.629	0.080	0.710	0.061
ϕ_{13}	0.799	0.088	0.794	0.084	0.818	0.057
ϕ_{14}	0.530	0.118	0.574	0.104	0.707	0.071
ϕ_{15}	0.700	0.137	0.716	0.115	0.794	0.071
ϕ_{16}	0.751	0.142	0.751	0.118	0.814	0.067
ϕ_{17}	0.859	0.124	0.820	0.112	0.840	0.065
ϕ_{18}	0.712	0.161	0.726	0.129	0.803	0.072
ϕ_{19}	0.832	0.139	0.802	0.118	0.832	0.068
ϕ_{20}	0.496	0.193	0.603	0.144	0.764	0.078
ϕ_{21}	0.751	0.191	0.751	0.145	0.816	0.072
ϕ_{22}	0.502	0.223	0.628	0.161	0.786	0.078
ϕ_{23}	0.680	0.236	0.721	0.156	0.813	0.075
ϕ_{24}	0.337	0.233	0.577	0.177	0.777	0.079
ϕ_{25}	0.497	0.294	0.670	0.175	0.804	0.076
ϕ_{26}	0.495	0.288	0.662	0.177	0.804	0.077
ϕ_{27}	0.490	0.289	0.660	0.176	0.805	0.077
ϕ_{28}	0.506	0.286	0.667	0.180	0.800	0.080
ϕ_{29}	0.494	0.293	0.664	0.177	0.799	0.079
λ	0.035	0.001	0.035	0.001	0.035	0.001

7.4 Convergence issues

We have not dealt explicitly in this paper with the problem of how to decide when the MCMC iterative procedures used reach equilibrium. This is discussed in Brooks *et al.* (2000a). Interestingly, we have found that convergence for the models that we have considered has been rapid, and better than expected. Typically, convergence was within 100-1000 iterations. This is most likely due to the constrained state-space imposed by the fact that the parameters of interest typically lie between zero and one. In practice, it is always important to check simulation performance and there are various simple-to-use convergence assessment techniques proposed in the literature (see for example, Brooks & Roberts, 1999).

An interesting potential of Bayesian analysis is that it can produce unique estimates when a likelihood surface has a flat ridge. An illustration of this is discussed in Brooks *et al.* (2000b), who explain that such a feature is sometimes associated with slow MCMC convergence. While this may be seen as an advantage

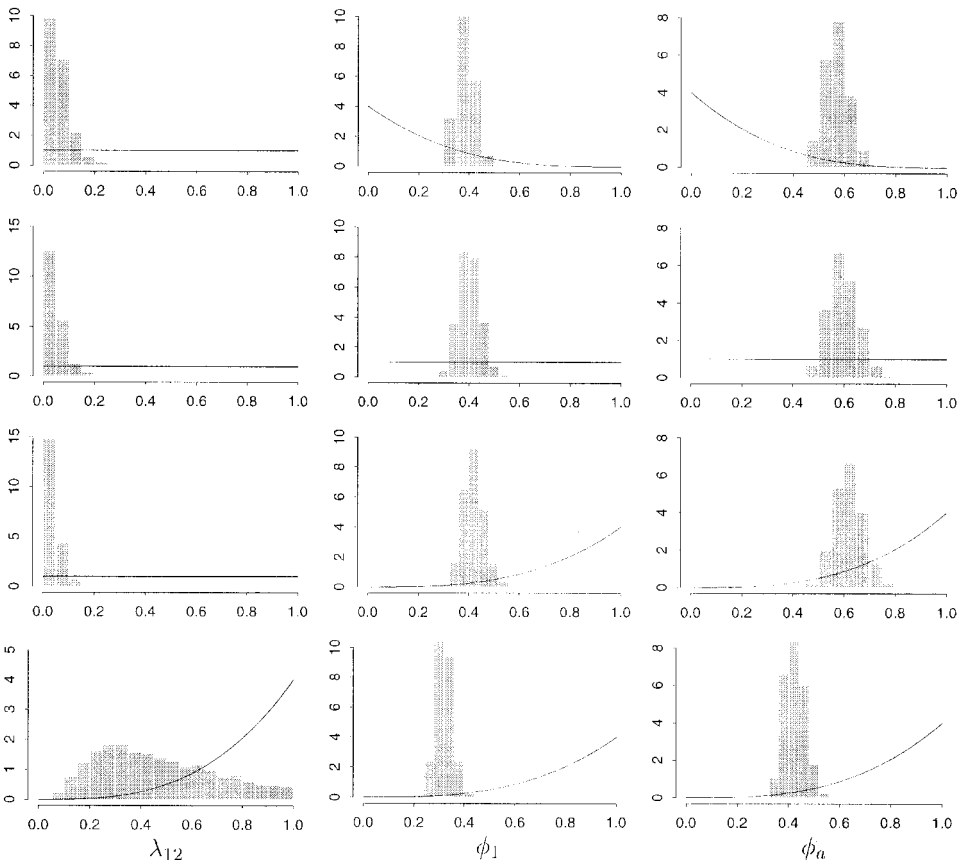


FIG. 4. Posterior histograms of λ_{12} , ϕ_1 and ϕ_a under the $\{\phi_1, \phi_a\}/\lambda_t$ model for the teal data, from a sample of 10 000 observations from the posterior distribution, with corresponding prior densities superimposed. The top row corresponds to beta(1,4) priors for the two ϕ parameters, the second row to Unif(0,1) and the bottom two rows to beta(4,1) priors; a Unif(0,1) prior is adopted for all the λ parameters in the first three rows and a beta(4,1) prior for the bottom row. From Brooks *et al.* (2000a).

of a Bayesian approach, circumventing the often thorny issue of parameter redundancy of models, it is also a potential pitfall if one does not realize the existence of the likelihood ridge.

8 Conclusions

We firmly believe that MCMC methods, especially RJMCMC approaches, will prove to be invaluable tools for the future analysis of mark recovery and recapture data. The work we have reported on using RJMCMC methods is ongoing, and we are still exploring its potential. We therefore agree with Schwartz & Seber (2000) that, ‘with the advent of the Gibbs sampler ... there will be an upsurge of interest in Bayesian methods ...’.

REFERENCES

ANDERSON, D. R. & BURNHAM, K. P. (1999) General strategies for the analysis of ringing data, *Bird Study*, 46 (Suppl.), pp. 261–270.

- BARRY, S. C., CATCHPOLE, E. A., BROOKS, S. P. & MORGAN, B. J. T. (2002) The analysis of ring-recovery data using random effects, *Biometrics*, submitted.
- BOLFARINE, H., LEITE, J. G. & RODRIGUEZ, J. (1992) On the estimation of the size of a finite and closed population, *Biometrical Journal*, 34, pp. 577-593.
- BROOKS, S. P. & ROBERTS, G. O. (1999) Assessing convergence of Markov chain Monte Carlo algorithms, *Statistics and Computing*, 8, pp. 319-335.
- BROOKS, S. P., CATCHPOLE, E. A. & MORGAN, B. J. T. (2002) Bayesian methods for analysing ringing data, *Journal of Applied Statistics*, this issue.
- BROOKS, S. P., CATCHPOLE, E. A. & MORGAN, B. J. T. (2000a) Bayesian animal survival estimation, *Statistical Science*, 15, pp. 357-376.
- BROOKS, S. P., CATCHPOLE, E. A., MORGAN, B. J. T. & BARRY, S. C. (2000b) On the Bayesian analysis of ring-recovery data, *Biometrics*, 56, pp. 951-956.
- BROWNIE, C., ANDERSON, D. R., BURNHAM, K. P. & ROBSON, D. S. (1985) *Statistical Inference from Band Recovery Data: a Handbook*, 2nd edn. US Fish and Wildlife Resource Publication 156.
- BUCKLAND, S. T., BURNHAM, K. P. & AUGUSTIN, N. H. (1997) Model selection: an integral part of inference, *Biometrics*, 53, pp. 603-618.
- BURNHAM, K. P. (1993) A theory for combined analysis of ring-recovery and recapture data. In: J.-D. LEBRETON & P. M. NORTH (Eds) *Marked Individuals in the Study of Bird Populations*, pp. 199-213 (Basel, Birkhäuser Verlag).
- BURNHAM, K. P. (2001) On random effects models for capture-recapture data, *J. Agric. Biol. Statist.*, submitted.
- BURNHAM, K. P. & ANDERSON, D. R. (1998) *Model Selection and Inference. A Practical Information-theoretic Approach* (New York, Springer).
- CARLIN, B. P. & LOUIS, T. A. (1996) *Bayes and Empirical Bayes Methods for Data Analysis* (London, Chapman & Hall).
- CASTLEDINE, B. (1981) A Bayesian analysis of multiple capture-recapture sampling for a closed population, *Biometrika*, 68, pp. 197-210.
- CATCHPOLE, E. A., FREEMAN, S. N., MORGAN, B. J. T. & HARRIS, M. P. (1998) Integrated recovery/recapture data analysis, *Biometrics* 54, pp. 33-46.
- CATCHPOLE, E. A., MORGAN, B. J. T., FREEMAN, S. N. & PEACH, W. J. (1999) Modelling the survival of British lapwings *Vanellus vanellus* using ring-recovery data and weather covariates, *Bird Study*, 46 (Suppl.), pp. 5-13.
- CATCHPOLE, E. A., MORGAN, B. J. T., COULSON, T. N., FREEMAN, S. N. & ALBON, S. D. (2000) Factors influencing Soay sheep survival, *Applied Statistics*, 49, pp. 453-472.
- CHATFIELD, C. (1995) Model uncertainty, data mining and statistical inference, *Journal of the Royal Statistical Society A*, 158, pp. 419-466.
- DI CICCIO, T. J., KASS, R. E., RAFTERY, A. & WASSERMAN, L. (1997) Computing Bayes factors by combining simulation and asymptotic approximation, *Journal of the American Statistical Association*, 92, pp. 903-915.
- DUPUIS, J. A. (1995) Bayesian estimation of movement and survival probabilities from capture-recapture data, *Biometrika*, 82, pp. 761-772.
- FREEMAN, S. N. (1990) *Statistical analysis of avian breeding and survival*, PhD Thesis, University of Kent, Unpublished.
- FREEMAN, S. N. & MORGAN, B. J. T. (1992) A modelling strategy for recovery data from birds ringed as nestlings, *Biometrics*, 48, pp. 217-236.
- GAMERMAN, D. (1997) *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference* (London, Chapman & Hall).
- GARTHWAITE, P. H., YU, K. & HOPE, P. B. (1995) Bayesian analysis of a multiple-recapture model, *Communications in Statistics*, 24, pp. 2229-2247.
- GEORGE, E. L. & ROBERT, C. P. (1992) Capture-recapture estimation via Gibbs sampling, *Biometrika*, 79, pp. 677-683.
- GILKS, W. R., RICHARDSON, S. & SPIEGELHALTER, D. J. (Eds) (1996) *Markov Chain Monte Carlo in Practice* (London, Chapman & Hall).
- GREEN, P. J. (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika*, 82, pp. 711-732.
- JANZ, R. J. (1980) Prior knowledge and ornithology. In: B. J. T. MORGAN & P. M. NORTH (Eds) *Statistics in Ornithology*, pp. 303-310 (Berlin, Springer-Verlag).
- KING, R. & BROOKS, S. P. (2001) Bayesian model discrimination for multiple strata capture-recapture data, *Biometrika*, 88, pp. 317-336.
- LEBRETON, J.-D., BURNHAM, K. P., CLOBERT, J. & ANDERSON, D. R. (1992) Modeling survival and

- testing biological hypotheses using marked animals: a unified approach with case studies, *Ecological Monographs*, 62, pp. 67-118.
- LEE, S. M. & CHEN, C. W. S. (2002) Bayesian inference of population size for behavioral response models, *Statistica Sinica*, to appear.
- LINDLEY, D. V. (1965) *Introduction to Probability and Statistics from a Bayesian View Point* (Cambridge University Press).
- LINK, W. A. & CAM, E. (2000) Of BUGS and birds: an introduction to Markov chain Monte Carlo, Unpublished ms.
- MADIGAN, D. & YORK, J. C. (1997) Bayesian methods for estimation of the size of a closed population, *Biometrika*, 84, pp. 19-31.
- MALAKOFF, D. (1999) Bayes offers a 'new' way to make sense of numbers, *Science*, 286, pp. 1460-1464.
- O'HAGAN, A. O. (1998) Eliciting expert beliefs in substantial practical applications, *The Statistician*, 47, pp. 21-36.
- PAYNTER, R. A. (1966) A new attempt to construct life tables for Kent Island herring gulls, *Bulletin of the Museum of Comparative Zoology at Harvard University*, 133, pp. 489-528.
- RACINE, A., GRIEVE, A. P., FLUHLER, H. & SMITH, A. F. M. (1986) Bayesian methods in practice: experiences in the pharmaceutical industry, *Applied Statistics*, 35, pp. 93-150.
- RAFTERY, A. E., MADIGAN, D. & HOETING, J. A. (1997) Bayesian model averaging for linear regression models, *Journal of the American Statistical Association*, 92, pp. 179-191.
- SCHWARTZ, C. J. & SEBER, G. A. F. (2000) A review of estimating animal abundance III, *Statistical Science*, 14, pp. 427-456.
- SMITH, A. F. M. & GELFAND, A. E. (1992) Bayesian statistics without tears: a sampling-resampling perspective, *The American Statistician*, 46, pp. 84-88.
- SMITH, P. J. (1988) Bayesian methods for multiple capture-recapture surveys, *Biometrics*, 44, pp. 1177-1189.
- SPIEGELHALTER, D. J., BEST, N. G. & CARLIN, B. P. (2001) Bayesian deviance, the effective number of parameters and the comparison of arbitrarily complex models, *Journal of the Royal Statistical Society, Series B*, in review.
- SPIEGELHALTER, D. J., THOMAS, A. & BEST, N. G. (1996a) Computation on Bayesian graphical models. In: J. M. BERNARDO, A. F. M. SMITH, A. P. DAWID & J. O. BERGER (Eds) *Bayesian Statistics 5*, pp. 407-425 (Oxford University Press).
- SPIEGELHALTER, D. J., THOMAS, A., BEST, N. G. & GILKS, W. R. (1996b) *BUGS: Bayesian Inference using Gibbs Sampling. Version 0.50* (MRC Biostatistics Unit, Cambridge).
- UNDERHILL, L. G. (1990) Bayesian estimation of the size of a closed population, *The Ring*, 13, pp. 235-254.
- VENABLES, W. N. & RIPLEY, B. D. (1999) *S Programming* (New York, Springer-Verlag).
- VOUNATSOU, P. & SMITH, A. F. M. (1995) Bayesian analysis of ring-recovery data via Markov chain Monte Carlo simulation, *Biometrics*, 51, pp. 687-708.
- WHITE, G. C. & BURNHAM, K. P. (1999) Program MARK: survival estimation from populations of marked animals. *Bird Study*, 46 (Suppl.), pp. 120-139.
- WILLIAMS, D. A. (1988) Estimating bias using the beta-binomial distribution in teratology, *Biometrics*, 44, pp. 305-307.
- ZEGER, S. L. & KARIM, M. R. (1991) Generalised linear models with random effects; a Gibbs sampling approach, *Journal of the American Statistical Association*, 86, pp. 79-86.

Appendix. BUGS code

The following BUGS code fits the random effects model (3) to the lapwing recovery data of Table 7.

```

model
{
  # Define the priors
  gam ~ dunif(0,25);
  sig ~ dunif(0,25);
  sigp <- 1/sig
  gamp <- 1/gam
  for (i in 1 : nj){
    psi[i] ~ dnorm(0, 0.625);
    b[i] ~ dnorm(0,sigp);
    for(j in 1: ni){
      e[j,i] ~ dnorm(0,gamp);
    }
  }
  lambda ~ dbeta(1, 1);

  # Define the likelihood
  for(i in 1 : ni){ m[i, 1 : nj + 1] ~ dmulti(p[i, ], r[i]); }

  # Calculate the no. of birds released each year
  for(i in 1 : ni){ r[i] <- sum(m[i, 1]); }

  # Random effects definition
  for(i in 1 : ni){
    for(j in 1 : (nj-i+1)){
      logit(phi[i,j]) <- psi[j]; # Model I
      logit(phi[i,j]) <- psi[j]+b[i+j-1]; # Model II
      logit(phi[i,j]) <- psi[j]+e[i,j]; # Model III
      logit(phi[i,j]) <- psi[j]+e[i,j]+b[i+j-1]; # Model IV
    }
  }

  # Calculate the cell probabilities
  for(i in 1 : ni){
    p[i, i] <- lambda * (1-phi[i,1]);
    for(j in (i+1) : nj){
      for(k in 1 : (j - i)){
        wp[i, j, k] <- log(phi[i,k]);
      }
      p[i,j] <- lambda*(1 - phi[i,j-i+1]) * exp(sum(wp[i,j,1:(j-i)]));
    }
    for(j in 1 : (i - 1)){
      p[i, j] <- 0; # lower triangle
    }
    p[i, nj + 1] <- 1 - sum(p[i, 1 : nj]) # Pr(never seen again)
  }
}

Data
list(ni = 23, nj = 23,
m= structure(.Data =
  c(8,3,2,0,0,0,1,2,3,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,1943,
    0,4,1,1,2,2,1,3,2,0,2,0,0,0,0,0,0,0,0,0,0,0,0,0,2445,
    0,0,7,2,2,2,5,1,5,2,1,3,1,1,1,2,0,0,0,0,0,0,0,0,3057,
    0,0,0,15,1,1,1,2,3,2,0,1,1,1,1,0,0,0,0,0,0,0,0,0,3413,
    0,0,0,0,12,4,4,7,4,1,1,1,1,0,1,2,1,1,0,0,0,0,0,0,3905,
    0,0,0,0,0,12,3,0,3,1,1,1,2,0,3,0,0,0,0,0,0,0,0,0,2538,
    0,0,0,0,0,0,11,3,5,1,3,4,2,3,0,1,0,1,1,0,0,0,0,0,3269,
    0,0,0,0,0,0,0,13,6,0,5,4,2,1,2,4,0,0,0,1,0,0,0,0,3440,
    0,0,0,0,0,0,0,0,18,5,2,2,0,5,3,0,0,0,1,0,0,0,0,0,3129,
    0,0,0,0,0,0,0,0,0,7,4,7,1,4,3,3,0,1,0,0,1,0,1,0,1,3319,
    0,0,0,0,0,0,0,0,0,0,12,8,2,2,4,5,3,0,1,2,0,0,1,1,3447,
    0,0,0,0,0,0,0,0,0,0,0,26,2,3,3,4,1,0,0,0,0,0,0,0,0,3897,
    0,0,0,0,0,0,0,0,0,0,0,0,11,0,6,3,0,1,2,0,0,1,0,2857,
    0,0,0,0,0,0,0,0,0,0,0,0,0,19,7,7,4,0,0,2,1,0,0,4079,
    0,0,0,0,0,0,0,0,0,0,0,0,0,13,3,2,0,0,0,0,1,0,4017,
    0,0,0,0,0,0,0,0,0,0,0,0,0,0,23,2,5,2,0,3,3,2,4827,
    0,0,0,0,0,0,0,0,0,0,0,0,0,0,13,4,3,3,4,2,2,4838,
    0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,14,2,1,2,3,2,5003,
    0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,17,4,4,3,0,4776,
    0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,11,4,2,4,3611,
    0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,12,3,3,4152,
    0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,9,4,4301,
    0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,18,3462
  ), .Dim = c(23, 24)))

Initial Values
list(lambda=0.5,
  psi=c(0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,
    0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5),
  sig=1.0,gam=1.0)

```